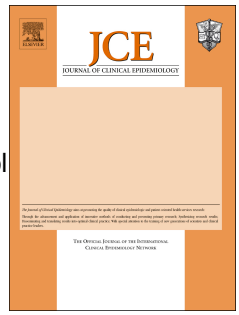
Accepted Manuscript



An overview of statistical methods for handling non-adherence to intervention protocol in randomised control trials (RCTs): A methodological review

Mohammod Mostazir, Rod S. Taylor, William Henley, Ed Watkins

PII: S0895-4356(18)30494-3

DOI: https://doi.org/10.1016/j.jclinepi.2018.12.002

Reference: JCE 9795

To appear in: *Journal of Clinical Epidemiology*

Received Date: 30 May 2018

Revised Date: 29 October 2018

Accepted Date: 4 December 2018

Please cite this article as: Mostazir M, Taylor RS, Henley W, Watkins E, An overview of statistical methods for handling non-adherence to intervention protocol in randomised control trials (RCTs): A methodological review, *Journal of Clinical Epidemiology* (2019), doi: https://doi.org/10.1016/j.jclinepi.2018.12.002.

An overview of statistical methods for handling non-adherence to intervention protocol in randomised control trials (RCTs): A methodological review

Mohammod Mostazir[a], Rod S Taylor[b], William Henley[b], Ed Watkins[c]

[a]College of life and Environmental Sciences (CLES), University of Exeter, EX4 4QG, United Kingdom, [b]Institute of Health Research, University of Exeter Medical School, University of Exeter, EX1 2LU, United Kingdom, [c]Mood Disorders Research, School of Psychology, College of Life & Environmental Sciences, University of Exeter, EX4 4QG, United Kingdom

Declarations of interest: None

Correspondence to:
Mohammod Mostazir
College of life and Environmental Sciences (CLES),
University of Exeter
EX4 4QG, Exeter UK
T: +44 1392 726629
M: +44 7432516771
Email: m.mostazir@exeter.ac.uk

Abstract

Objective

To undertake a methodological review of statistical methods used in randomised controlled trials (RCTs) for handling intervention non-adherence.

Study design

Bibliographic databases were searched using predefined search terms..

Results

A substantive number of identified studies (56%) were excluded as they only used naive per-protocol (PP) analysis for handling non-adherence. Our review included 58 articles published between 1991 to 2015. A total of 88 methodological applications were made by these studies. The two most used methods were Complier Average Causal Effect (CACE) (56%) and Instrumental Variable (IV) (23%) predominantly with the use of maximum Likelihood (ML) estimators. These alternative applications typically produced treatment effects greater than the intention-to-treat (ITT) effect but as their standard errors were larger there was no statistical difference between the methods.

Conclusion

A substantive proportion of RCTs rely on naive PP for handling non-adherence. Recent years have seen an increasing number of applications of more appropriate statistical methods, in particular CACE and IV methods. However, these later methods rely on strong underlying assumptions that may be vulnerable to violation. More empirical studies are needed that directly compare the usability and performance of different statistical methods for non-adherence in RCTs.

Keywords: Non-adherence; Non-compliance; Randomised controlled trial; Methodological review; Causal effect modelling; Statistical methods

Running title: Methodological review of statistical methods for handling non-adherence in RCTs.

(Word count: 192, excluding Title, Headings, Keywords and Running title.)

What is already known

- Randomised controls trials (RCTs) often suffer from non-adherence or non-compliance of trial participants to the intervention(s) protocol to which they are randomised.
- Per-protocol (PP) and as-Treated (AT) are two naïve analytical methods for handling non-adherence in RCTs, which are prone to serious selection bias and cannot claim causal treatment effect.
- Several statistical applications based on causal inference are now available to more appropriately adjust treatment effect for non-adherence in RCTs data.

What this study adds

- Our methodological review shows that, a large proportion of RCTs continue to rely on naïve PP method for handling intervention non-adherence.
- Maximum likelihood (ML) based Complier Average Causal Effect (CACE) and Instrumental Variable (IV) are more appropriate approaches to handling non-adherence in RCTs.

(Word count: 3,054 excluding headings, subheadings, tables/figures, conflict of interest, authors' contribution, references)

1. Introduction

Randomised controlled trials (RCTs) and systematic reviews of RCTs provide the highest level of evidence for assessing the effects of healthcare interventions.[1] Researchers, however, still face challenges when undertaking RCTs. One of these is the non-adherence/non-compliance of trial participants to the intervention(s) protocol to which they are randomised.

Non-adherence has been shown to be associated with poorer patient outcomes, including higher mortality.[2] A meta-analysis across 569 trials estimated an average treatment non-adherence rate of 25%,[3] while another study reported the rate 23%.[4] Current reporting guidelines for RCT (Consolidated Standards of Reporting Trials [CONSORT]) recommend the intention-to-treat (ITT) approach i.e. outcomes are compared according to original group allocation regardless of whether participants received the intervention according to the protocol or not.[5, 6] By doing so, ITT evaluates effectiveness of an intervention by mirroring the non-adherence to treatment that may occur in real-world practice. Whilst this may be true, it is argued that by ignoring non-adherence, ITT underestimates the 'true (or causal) effect' of the intervention because the analysis is diluted by non-compliers.[7-9]

A commonly used approach by analysts to handling non-adherence is per protocol (PP) analysis where the outcomes of intervention are compared according to initial random allocation but excluding those participants who do not adhere to the intervention protocol.[10, 11] A systematic review of 100 RCTs identified 47% studies to have adopted some form of PP analysis.[12] The PP approach is prone to serious selection bias as it fails to preserve the original randomisation and causality of treatment effect cannot be claimed.[13] 'As treated' (AT) analysis is another variant of non-ITT analysis,[14, 15] that classifies participants according to the intervention they receive regardless to their adherence to the trial protocol and like PP analysis is subject to selection bias.[16, 17]

Several statistical methods have been developed for estimating causal treatment effects that take account of intervention non-adherence without introducing the biases inherent to PP or AT analyses. The statistical framework for causal inference in RCTs was

developed by Rubin, referred to as Rubin's Causal Model (RCM) where each participant is assumed to have a set of counterfactual outcomes.[18-21] Under the RCM framework, several methods developed for handling non-adherence, including Instrumental Variable (IV) approach from the field of econometrics,[22] Complier Average Causal Effect (CACE) by Rubin,[14, 23] and Structural mean models (SMM) by Robins.[24] To our knowledge there has been no comprehensive review of the use of these statistical methods and their pros and cons.

We undertook a methodological review of RCTs that described statistical methods for handling non-adherence to intervention protocol. Given the bias associated with the methodology, we excluded studies that utilised PP analysis alone. The aims were to: (1) assess the range of statistical methods reviewed and applied in RCTs to handle non-adherence; (2) review the relative pros and cons of these statistical methods; (3) Make a pooled comparison of the treatment effects estimated by ITT and proposed statistical methods for handling non-adherence.

2. Methods

We conducted and reported this methodological review in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement.[25]

2.1 Literature search strategy

We searched a number of bibliographic databases i.e. EMBASE (OvidSP), PsycInfo (OvidSP), MEDLINE (OvidSP), CINAHL (EBSCOHOST) and Cochrane Library for Methodological Studies (Wiley Online Cochrane Library) from inception to June 2015. Database specific Boolean search strategies were developed using key terms i.e. 'intention to treat', 'as-treated', 'per protocol', 'non-adherence', 'complier average causal effect', 'CACE' (and synonyms). The reference lists of the included papers were manually checked. Details of the search strategy are provided in the e-appendix (A).

2.2 Study selection

We included RCTs that reviewed statistical methods for handling non-adherence and applied these methods to actual/simulated trial participant data. Studies were excluded if: (1) they were available only as abstracts/titles and not as a full publication; (2) they adjusted for non-adherence but provided no information on the statistical basis of this method (this included studies that simply stated that they used 'IV' or 'CACE' analysis but gave no further

methodological details);[26] (3) they applied statistical methods for handling any potential confounding/bias but this was unrelated to non-adherence to intervention protocol.[27]

2.3 Data extraction

A database was compiled that captured information on characteristics of included studies, i.e. title, authors, journal, year of publication, population disease area, type of intervention, randomizing unit, study duration, type of outcomes, sample size and estimated treatment effect by ITT method and by the proposed methods. Detailed information was extracted on the method of statistical analysis applied i.e. name of the statistical method/framework, statistical estimators/algorithm applied to implement the technique and any advantages/disadvantages of these statistical method as stated by authors.

2.4 Data analysis and presentation

A descriptive approach was taken to data presentation using frequencies, means and medians. Pooled comparison of direct treatment effects across studies was not feasible as studies had varied outcomes. For comparison of treatment effect between ITT and the proposed methods, we compared whether the treatment effect by ITT was larger or smaller compared to the effect estimated by the proposed method (coded 'yes/no') and presented the results in frequency (%). Further, absolute z-statistic [(treatment effect / standard error (s.e.)] was calculated for each method application and the pooled mean z-statistic was compared between ITT and proposed methods. This pooled comparison accounted for within-study variance by subtracting each proposed method z-statistic from ITT-z-statistic before calculating pooled mean z-statistic i.e. $\sum [(Z_p - Z_{ITT}) / n]$, where Z is the z-statistic for proposed (p) or ITT method and n is the number of applications. The pooled z-statistic was also used to compare treatment effect between IV vs CACE method by meta-regression accounting for non-adherence rate. Applications made on simulated data and applications involving Bayesian method were excluded from these comparisons. Authors presented information in various formats i.e. presenting coefficient and 95% confidence interval (CI), presenting coefficient and s.e or presenting coefficient and the P-value only. We derived required statistic applying appropriate formulae,[28-30] where applicable. Where results from several models were presented i.e. model comparisons from sensitivity analyses, for data extraction, we considered the optimum model suggested by the author. Analyses were undertaken using statistical software Stata, version 15.[31]

3. Results

3.1 Selection of included RCTs

The literature searches resulted in a total of 4,664 titles/abstracts, of which 58 were eligible for inclusion (Figure 1). A total of 2,591 (56%) of the abstracts were excluded because authors exclusively relied on PP analysis to deal non-adherence. A small number of studies were also excluded for applying AT analysis (26 studies, 0.56%) and modified ITT (84 studies, 1.8%) as both are forms of PP analysis.[12] The other reasons for exclusions were applications that were non-RCTs, unrelated to handling non-adherence or lacked methodological details.

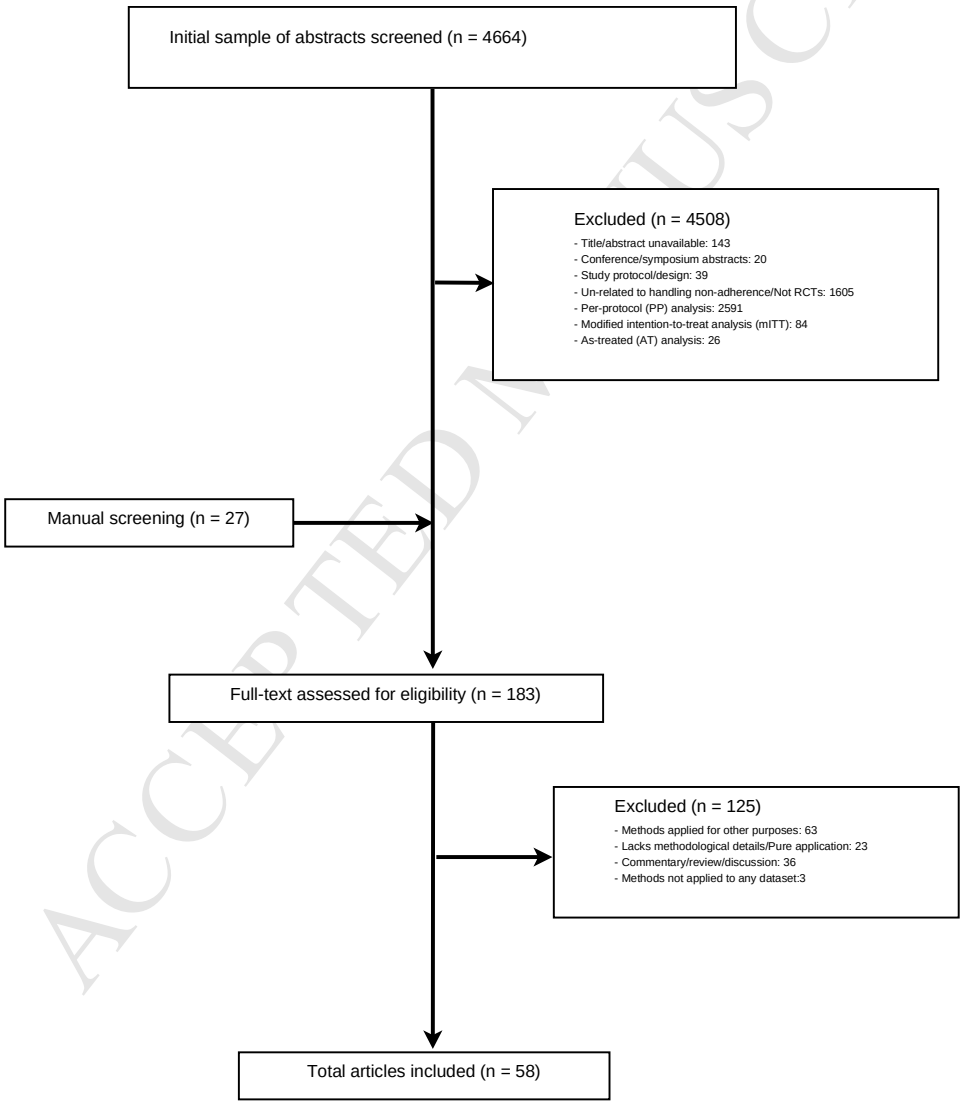


Figure 1: Flow of studies through inclusion and exclusion process

3.2 Characteristics of included RCTs

Detailed description of included studies is given in the e-Appendix (B). Summary of study characteristics is presented in Table-1. Majority of included studies were published in statistical/methodological journals. Studies were undertaken across a wide range of patient and intervention types, study sizes, duration and were applied across a range of outcome types (continuous/binary/count/time-to-event).

Characteristics	Number	Percent (%)
Number of articles	58	100
Year of publication (n = 58)	–	–
1991-1999	12	21
2000 - 2007	21	36
2008 - 2015	25	43
Journals (n = 58)	–	–
Statistics in medicine	20	34
Biometrics	8	14
Biostatistics	5	9
Journal of the American Statistical Association	4	7
Controlled Clinical Trials	2	3
Journal of the Royal Statistical Society	2	3
Psychological Methods	2	3
American journal of epidemiology	1	2
Biometrical Journal	1	2
Biometrika	1	2
British Journal of Psychiatry	1	2
Clinical trials	1	2
Family process	1	2
Health Services & Outcomes Research Methodology	1	2
Journal of Biopharmaceutical Statistics	1	2
Journal of Clinical Epidemiology	1	2
Journal of Educational and Behavioral Statistics	1	2
Psychological Medicine	1	2
Statistica Sinica	1	2
Statistical Methods in Medical Research	1	2
The American journal of drug and alcohol abuse	1	2
The annals of statistics	1	2
Study population clinical area (n = 58)	–	–
Mental health	17	29
Cardiology	7	12
Infectious disease	5	9
Oncology	4	7
Others	20	34
Simulation study/NA	5	9
Follow up duration (n = 58)	–	–
1 year	11	19
1-2 year	19	33
>2 years	15	26
Not reported	13	22
Type of intervention (n = 58)	–	–
Drug	21	36
Psycho-therapy	8	14
Behavioural	8	14
Other	16	28
Simulation study (N/A)	5	9
Type of outcome (n = 58)	–	–
Binary	29	50
Continuous	19	33

Count/Time-to-event	10	17
Sample size (n = 58)	–	–
<400	19	33
400-1000	14	24
>1000	18	31
Not-reported/NA	7	12
Randomising unit (n = 58)	–	–
Individual	46	79
Cluster	5	9
Meta-analysis	1	2
Simulation study (N/A)	6	10
% Non-adherence: Median (range)	–	34% (2% to 78%)

Table 1: Summary of included study characteristics

A significant rise was observed in the published literature in this area of methodological research since 1999. Per-year average rate of publication (IRR: incidence rate ratio) was higher at the later periods compared to the 1991-1999 period (IRR for 2000-2007: 1.56, 95% CI: 1.00 to 2.45, $p<0.05$; IRR for 2008-2015: 1.72, 95% CI: 1.11 to 2.65, $p<0.01$).

3.3 Statistical methods and estimators used in included RCTs to handle non-adherence

A total of nine methods for handling treatment non-adherence were described across the included studies (Table 2a). Some of these studies applied more than one method using different estimators resulting in a total of 88 statistical method applications. Studies that were judged to be variants of a common statistical approach were grouped under broader approach i.e. 'Longitudinal Complier Average Causal Effect' or 'Complier Average Causal Effect within Effect Class (ECACE)' were grouped under 'Complier Average Causal Method (CACE)' as they broadly have similar statistical framework. The two most common statistical methods applied were the Complier Average Causal Effect (CACE) (49/88 applications, 56%) and Instrumental Variable (IV) (20/88 applications, 23%). Authors applied the term 'IV' for both 'method' and 'estimator' where 'IV-method' refers to the causal instrumental variable framework[23] and 'IV-estimator' refers to implementation of a particular method applying two-stage least square (2SLS) estimator.[32] A total of 10 estimators (Table 2b) were identified, the most common being Maximum Likelihood (ML) estimator in 33% (n = 29), Method of Moment (MOM) base estimator in 22% (n = 19), Instrumental Variable (IV)

estimator in 19% (n = 17) and Bayesian estimators were used in 10% (n = 9) of the applications.

Sl #	a) Methods	Method elaboration	Number (%)	b) Estimators	Estimator elaboration	Number (%)
1	**CACE**	Complier average causal effect model	49 (56)	**ML**	Maximum likelihood	29 (33)
2	**IV**	Instrumental variable model	20 (23)	**MOM**	Method of moments	19 (22)
3	**SNMM**	Structural nested mean model	7 (8)	**IV**	Instrumental variable estimator	17 (19)
4	**ATR**	Adjusted treatment received model	4 (5)	**BI**	Bayesian inference	9 (10)
5	**RPSFTM**	Rank preserving structural failure time model	3 (3)	**G-estimator**	G-estimator	5 (6)
6	**C-PROPHET**	Rank preserving structural failure time model	2 (2)	**Cox-PH**	Cox-proportional hazard estimator	4 (5)
7	**CALM**	Compliers proportional hazards effect of treatment with proportional Hazards model	1 (1)	**WLS**	Weighted least square	2 (2)
8	**Cox-Reg1**	Causal accelerated life model	1 (1)	**GSMM**	Generalized structural mean model estimator	1 (1)
9	**Cox-Reg2**	Regression adjustment with Cox-model	1 (1)	**ISM**	Intensity score method	1 (1)
				WGSNM	Weighted generalized structural mean model	1 (1)
	Total		**88 (100)**			**88 (100)**

Table 2: Statistical methods (a) and their estimators (b) as stated by the authors, applied for handling non-adherence

Figure 2/a shows the different estimators used for applications of CACE methods. ML base estimators were implemented with Expectation Maximization (EM) algorithm,[33] and Bayesian Inference (BI) base methods were implemented both with EM and Markov-Chain Monte-Carlo (MCMC) algorithm.[34] As shown in Figure 2/a, 49% (n = 24) of the CACE applications were made using ML estimators. Figure 2/b shows all other statistical methods and the frequencies of the use of their different estimators.

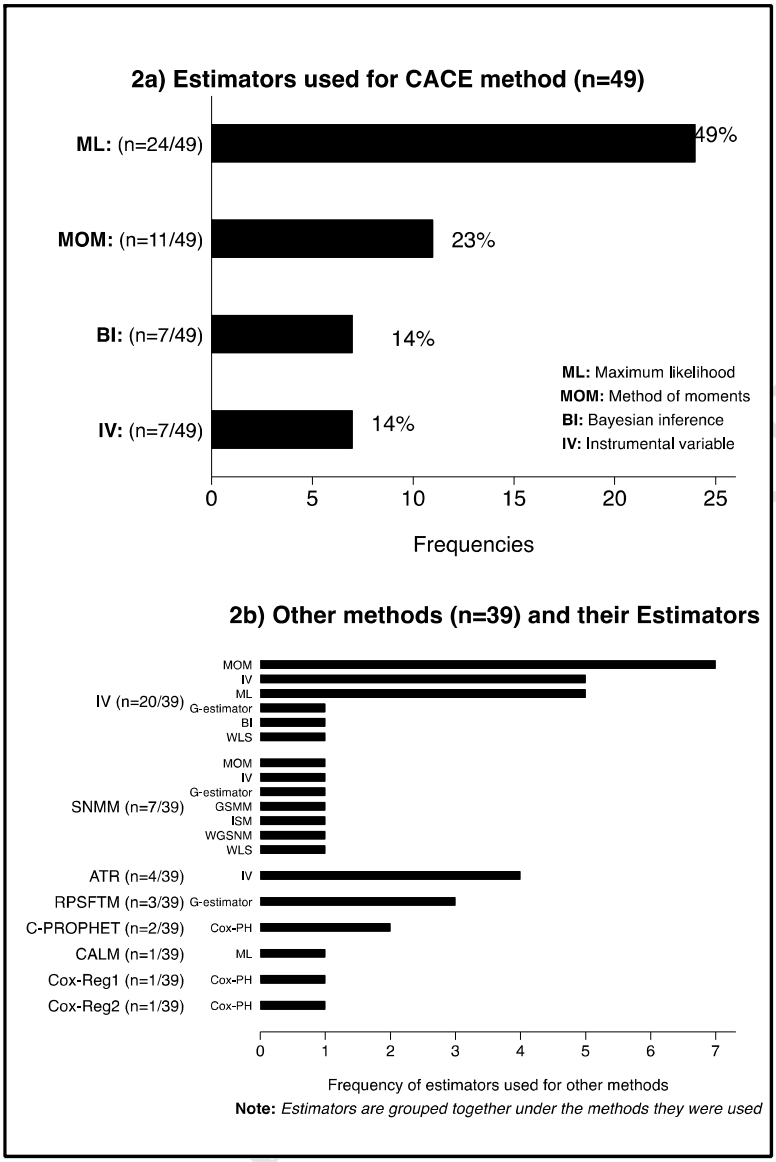


Figure 2: a) Use of varied estimators for estimating CACE and (b) other methods

3.4 Pros and cons of statistical methods presented by authors

The remainder of the methods section and Table 3 provide an overview of the statistical basis of the statistical methods and stated pros and cons of these approaches.

3.4.1 Complier Average Causal Effect (CACE)

Based on counterfactual outcome,[21] the CACE method was introduced by Angrist et al. for estimating causal effects in the presence of non-adherence.[23] In CACE analysis the potential adherence classes are stratified into four principal strata based on principal stratification,[35] i.e. i) 'Compliers' i.e. receive treatment when they are assigned to it, ii) 'Never-takers' i.e. do not receive treatment when they are assigned to it, iii) 'Always-Takers'

i.e. always receive the treatment regardless of randomisation and iv) 'Defiers' i.e. always do the opposite of what is assigned and assumed to be non-existent. In addition to randomisation and the stable unit treatment value assumption (SUTVA),[36, 37] there are two key assumptions that need to be fulfilled for CACE model to be identified: (1) the effect of treatment assignment on outcomes entirely operates through treatment receipt status of participants, known as "exclusion restriction" (ER). ER in other words states that under true randomization, the proportion of non-compliers in the control group (had they been offered the treatment) and their outcomes are similar to the proportion of observed non-compliers and their outcomes in the treatment group; (2) The "monotonicity" assumption implies that there are no 'defiers' meaning no participants will refuse treatment when assigned to treatment and will seek treatment when assigned to control. Though the initial CACE estimator proposed by Angrist et al. was an IV estimator,[23] we identified several other estimators for CACE applied into different settings (Figure 2a). Our findings, across several types of CACE applications, suggest that ML-base estimation was applied more often than other estimators and the reason may be that ML estimates are considered more efficient than 2SLS (Two-Stage Least Square) based IV estimators.[38-40] We also found applications of IV estimators in combination with ML estimators in estimating CACE and this combination contributed to substantial methodological development.[41] Missing data adds another level of complexity in presence of treatment non-adherence and in over half of CACE applications (24/47) authors provided guidelines for handling missing data. CACE also has been implemented in cluster randomised trials where intra-class correlation (ICC) from similar adherence behaviour at cluster level may compromise estimated treatment effects.[42-45] When there are multiple arms involved, CACE model may suffer from non-identifiability issues or may require complex modelling assumptions,[46, 47] and Bayesian methods may be applied addressing such complexities.[48] The fundamental limitation of the CACE approach is that the underlying assumptions i.e. ER, monotonicity are not easily testable,[49-52] and if violated CACE estimates may be biased.[41, 52, 53]

3.4.2 Instrumental variable (IV)

An IV is an exogenous variable that influences the outcome solely through a binary post-treatment variable that identifies whether participants adhered to treatment or not.[21, 22] Typically in RCTs, an IV is the randomizing variable and participants' adherence status is the endogenous variable through which outcome is affected. The assumption that outcome solely depends on adherence status is equivalent to the ER assumption discussed in the CACE section above. Therefore, in a two-arm trial design where participants' choice to post-randomisation switching between arms is restricted, an IV estimates alternate CACE

estimates given same estimator applied.[49] Typically IV estimators are implemented with 2SLS[32] estimators, however, ML-based IV estimators are also used.[54] In our selected studies, we separately identified ML-base IV estimators where authors were explicit about it. We found IV methods, like CACE, being applied in varied scenarios. However, IV with 2SLS is likely to estimate treatment effect on complete case basis and valid only when missing data are ignorable.[32] When compliance rate is low, 2SLS-base IV estimator produces large effects compared to ITT and produces large variances which makes it a less attractive estimator.[55] In such scenarios, ML is a more efficient estimator of IV.[56] A variation of IV method is Adjusted Treatment Received (ATR) method introduced by Nagelkerke[57] with an adjustment made to error terms. The distinction to typical IV method is that in ATR, the error terms from first stage endogenous regression is added to the model as a covariate to allow adjustment for any unmeasured confounding.

3.4.3 Other statistical methods

Structural Mean Model (SMM)/Structural Nested Mean Model (SNMM) was introduced by Robins.[24] The framework provides causal treatment effect for observed adherence comparing with a conditional reference level of adherence.[47] Linear additive framework is used for continuous outcomes and multiplicative framework is used for binary outcomes. Models are estimated with the G-estimator (GE) proposed by Robins and Tsiatis.[58, 59] The appealing aspect of SMM is that causal parameters can be estimated for varying levels of adherence. However identifying reference level of compliance may be challenging.[60] Another version of SMM applied to accelerated failure time (AFT) model (time to event survival data with time as outcome) is Rank Preservative Structural Failure Time model (RPSFTM).[61, 62] They are called rank preserving because they use a class of rank estimators for subjects' failure.[62] In practice, G-estimators have not been widely adopted due to level of complexities involved in implementation.[63] For handling non-adherence in continuous time survival data include Cox-reg$_{(1,2)}$, Complier proportional hazard effect of treatment (C-PROPHET) model and Causal accelerated life model (CALM).[64, 65] Cox-reg$_{(1,2)}$ both are adherence adjustment base Cox-regression implemented by Cox-PH estimator. Cox-reg$_1$ is implemented in situations when compliance is 'all-or-nothing' i.e. either patients receive the treatment, or they do not and Cox-reg$_2$ is implemented when compliance is partial. CALM and C-PROPHET both have CACE like framework where CALM is applied to AFT model and C-PROPHET is applied to continuous time survival data.

Methods*	Frequency	Estimators*	Method description	Strengths	Limitations in implementation
CACE	49	BI/IV/ML/MOM	Based on principal stratification and counterfactual outcome rather observed outcome, CACE estimate relates to treatment effect for participants who would have complied with the treatment had they been offered it.	i) In presence of non-adherence, unlike PP/AT, it yields causal effect of treatment on the treated ii) Randomization based estimate of efficacy iii) By using pre-treatment covariates identifiable CACE model is feasible instead forcing exclusion restriction assumption. iv) Cell specific i.e. 'always-takers', 'never-takers' treatment effect estimation is possible.	i) Assumptions are not directly testable and may not reflect the real world scenario under non-adherence ii) CACE estimates can be biased if assumption is violated and the bias can be substantial with low compliance rate iii) The strong assumptions limits the flexibility of CACE modelling in practice iv) Encounters difficulty when number of randomized arms and adherence categories increase.
IV	20	BI/G-estimator/IV/ML/MOM	Instrumental variable method historically grounded in econometric theory where an instrument is an exogenous variable that influences the outcome through adherence related post-treatment variable only.	i) IV estimate does not require an assumption of homogeneous treatment effects, under exclusion restriction and the monotonicity assumptions ii) Method is not sensitive to differences in baseline risk between compliers and non-compliers iii) IV tends to have superior RMSE unless the compliance rate is low or zero.	i) Outcome may be affected by other means rather solely through treatment received and randomization. ii) Method is most sensitive to violations of exclusion restriction and to the monotonicity assumption when there are few compliers. iii) With 2SLS estimator and low compliance rate, IV can produce large variances iv) Implementation with 2SLS can only be feasible if missing data are ignorable.
SNMM	7	GE/WGSNM/ISM/IV/MOM/WGSNM/WLS	Method relates to comparing mean of the outcome at an observed compliance level with the mean of the potential outcome at some reference level.	i) Provide randomization based causal effects at varying level of adherence ii) Cause–effect relationships are established by considering a potential treatment-free outcome for each observational unit.	i) With binary outcomes, in some instances with complex study design i.e. three arms, the estimating equation has no solution. ii) G-estimation can be complex in implementation.
ATR	4	IV	Variation of IV method with error term from endogenous regression added in the model as covariate	i) Method copes with missing data problems as the first stage uses all randomized participants. Only the second stage is affected with missing data	i) Method only valid when patients switch between treatment arms, for example when one arm consists of placebo therapy and a placebo effect is not anticipated.
RPSFTM	3	GE	Estimates parameters of a class of semi-parametric failure time models, using a class of rank estimators. These models are the structural version of the "accelerated failure time model with time-dependent covariates"	i) Yields valid results for both outcome-dependent and outcome-independent treatment non-compliance ii) designed to consistently estimate causal effects on the treated, without direct assumptions about the compliance selection mechanism	i) RPSFTM makes a strong non-interaction assumption which in certain settings might be considered biologically implausible ii) G-estimation can be complex in implementation.
C-PROPHET	2	Cox-PH	Randomization based proportional hazard model for continuous survival data	i) Model framework is similar to CACE and provides estimates equivalent to CACE	i) Allows only binary/all or none compliance ii) I Under time-varying noncompliance the model maybe biased
Cox-Reg$_{1-2}$	2	Cox-PH	Cox-Proportional Hazard model for continuous survival data adjusting for non-adherence	i) Yield valid causal estimates under random non-adherence	i) Under non-random adherence, estimates can be biased
CALM	1	ML	Accelerated Failure to Time (AFT) model for survival data that relates each observed event time in the treated group to a potential event time that would have been observed if the control treatment had been given throughout the trial	i) Allows cross-over type non-adherence between arms ii) Estimates hazard ration equivalent to CACE ii) Does not rely on homogenous population	i) The model may produce extreme values and larger error terms

*Please refer to Table 2 for the elaboration of the method/estimator acronyms

Table 3: Summary of methods, pros and cons as stated by the authors

3.5 Comparison of estimated treatment effects

We were able to compare treatment effect for 68/88 applications. The majority of the alternative methods (n = 48, 71%) produced treatment effects that were greater than the treatment effect estimated by ITT. For 11 applications (16%) estimates were similar for both ITT and the proposed methods. For all alternative methods, excluding the Bayesian applications, 95% CIs overlapped with the CIs of ITT either at lower or upper bound region.

64/88 applications contributed to the calculation of standard errors and z-statistics. In 83% of the applications (53/64), standard errors for the alternative methods were larger than the s.e. of ITT estimates. After accounting for within study variation, average z-statistic from proposed methods were greater by +0.13 SD (95% CI: -0.99 to 1.71). We found 7 out of 58 studies (12%) achieved significant treatment effect by applying an alternative method which was not achieved by the ITT method.

In meta-regression, when accounted for percent non-adherence rate, z-statistic for IV method was no different than z-statistic from ITT (-0.01, 95% CI: -0.27 ti 0.26) but z-statistic from CACE were greater by +0.18 SD (0.18, 95% CI: -0.01 to 0.35). CACE estimates were higher by the same amount when compared to IV.

4. Discussion

In this review, across 58 studies, a wide variety of statistical methods against ITT were identified for handling treatment non-adherence. The median intervention non-adherence was 38% ranging across studies from 2% to 78%. The two most commonly used methods were Complier Average Causal Effect (CACE) and Instrumental Variable (IV). Overall, there was no significant difference between the pooled z-statistics from ITT and the alternative methods. In general, the majority of the proposed applications (83%) produced larger error variance compared to the error variance produced by ITT. We note that use of the CACE method resulted in larger z-statistics compared to the IV method when accounting for non-adherence rate.

We are aware of two previous systematic reviews undertaken to assess the analytical approaches to the handling of treatment protocol non-adherence in RCTs. Dodd et al.[12] summarised the extent to which non-adherence to treatment protocol is reported in RCTs. However, this study did not identify methods apart from the conventional ITT, PP and

AT analysis approaches. Adewuyi et al.[66] studied non-adherence in surgical intervention and reported that 63% of the studies adopted ITT, 21% PP and 3% AT analysis. Our systematic review is therefore the first to identify and systematically reviewed statistical methods that have been developed to handle non-adherence using a causal inference framework.

The CACE and IV methods are flexible and have been applied across a range of RCT designs. One of the benefits of the CACE application is that cell-specific treatment effect can be obtained which can provide valuable insights for researchers in relation to various types of adherence, whereas this opportunity is limited for the IV approach. We also found good number of applications of MOM estimators (19/88, 22% applications), but we avoided emphasizing on it as it relies on simple cell means and ignores distributional error terms which can be erroneous.[52, 67] According to our findings, both CACE and IV methods are applicable to varieties of scenarios and both rely on strong assumptions that are vulnerable to violations. Unless there are direct ways of testing the assumptions, it is not readily verifiable whether the applied methods captured the true treatment effect or simply inflated the treatment effect influenced by level of adherence.

5. Strengths and limitations

The strength of this study was its use of a systematic review approach to identify studies for inclusion. However, it has a number of limitations. Study selection and data extraction was undertaken by a single reviewer (MM) although the opinion of a second reviewer (RST) was available. We were unable to compare treatment effects estimated by different statistical methods because of their varied outcomes. The comparison of pooled z-statistic may not be an ideal approach, but this provides an indication of location of treatment effect estimated by different methods around the region of significance. We excluded studies that implemented relevant methods purely for application purposes instead of providing methodological guidance for handling sub-optimal adherence and also studies that used statistical methods for handling general confounding other than handling non-adherence exclusively e.g. propensity score (PS), inverse probability weighting (IPW). These methods have wider applications in observational studies for adjusting general confounding based on probabilistic weighting, but they directly do not contribute to the formation of causal frameworks for handling non-adherence.

6. Implications for practice and policy

Usually the ITT estimate of a treatment effect will be smaller than the ‘true’ effect since if the treatment works, non-compliance to treatment means suboptimal effects.

Therefore, the search for an alternative to ITT is a growing area of interest. Perhaps surprisingly, we found that a large number of RCTs continue to use PP methods despite the major limitations of this approach. CACE and IV methods are two important unbiased alternatives to ITT when adherence to treatment is sub-optimal and this review shows that these methods have been applied to a wide range of RCTs. However, given both suffer from strong underlying assumptions these methods are always reported in addition to ITT analysis and regarded as a sensitivity analysis.

7. Conclusions

Our review found that the alternative methods for handling non-adherence rely on strong assumptions that may be vulnerable to violations. More empirical studies are needed that directly compare the usability and performance of different statistical methods for non-adherence in RCTs.

Conflict of interest

None declared.

Authors' contributions

MM undertaken literature search, carried out data extraction, statistical analysis, and drafting the manuscript. RST reviewed article selection/data extraction process and provided guidance on overall structure of the paper. WH reviewed the statistical aspects of the paper and EW provided constructive inputs in article structure, formats and presentation of results. All authors were involved in writing the paper and had final approval of the submitted and published versions.

References

1. Charlton BG. Medical practice and the double-blind, randomized controlled trial. The British Journal of General Practice. 1991;41(350):355-6.

2. Simpson SH, Eurich DT, Majumdar SR, et al. A meta-analysis of the association between adherence to drug therapy and mortality. BMJ. 2006;333(7557):15.

3. DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. Med Care. 2004;42(3):200-9.

4. Bannister-Tyrrell M, Miladinovic B, Roberts CL, Ford JB. Adjustment for compliance behavior in trials of epidural analgesia in labor using instrumental variable meta-analysis. Journal of Clinical Epidemiology. 2015;68(5):525-33.

5. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ. 1999;319(7211):670-4.

6. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340.

7. Hernán MA, Hernández-Díaz S. Beyond the intention to treat in comparative effectiveness research. Clinical trials (London, England). 2012;9(1):48-55.

8. Gupta SK. Intention-to-treat concept: A review. Perspectives in Clinical Research. 2011;2(3):109-12.

9. Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. Clinical pharmacology and therapeutics. 1995;57(1):6-15.

10. Sedgwick P. Per protocol analysis. BMJ. 2010;340.

11. Sedgwick P. What is per protocol analysis? BMJ : British Medical Journal. 2013;346.

12. Dodd S, White IR, Williamson P. Nonadherence to treatment protocol in published randomised controlled trials: a review. Trials. 2012;13(84).

13. Sedgwick P. Intention to treat analysis versus per protocol analysis of trial data. BMJ (Clinical research ed). 2015;350:h681.

14. Frangakis C, Rubin D. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. Biometrika. 1999;86(2):365-79.

15. Montori VM, Guyatt GH. Intention-to-treat principle. CMAJ. 2001;165(10):1339-41.

16. Ten Have TR, Normand S-LT, Marcus SM, et al. Intent-to-treat vs. non-intent-to-treat analyses under treatment non-adherence in mental health randomized trials. Psychiatric Annals. 2008;38(12):772-83.

17. Bang H, Davis CE. On estimating treatment effects under non-compliance in randomized clinical trials: are intent-to-treat or instrumental variables analyses perfect solutions? Stat Med. 2007;26(5):954-64.

18. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology. 1974;66(5):688-701.

19. Rubin DB. Bayesian Inference for Causal Effects: The Role of Randomization. 1978:34-58.

20. Holland PW. Statistics and Causal Inference. Journal of the American Statistical Association. 1986;81(396):945-60.

21. Marcus SM, Gibbons RD. Estimating the efficacy of receiving treatment in randomized clinical trials with noncompliance. Health Services and Outcomes Research Methodology. 2001;2(3-4):247-58.

22. Bloom HS. Accounting for No-Shows in Experimental Evaluation Designs. Evaluation Review. 1984;8(2):225-46.

23. Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. Journal of the American Statistical Association. 1996;91(434):444-55.

24. Robins J. Correcting for non-compliance in randomized trials using structural nested mean models. Communications In Statistics : Theory And Methods. 1994;23(8):2379-412.

25. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097.

26. Osypuk TL, Tchetgen Tchetgen EJ, Acevedo-Garcia D, et al. Differential mental health effects of neighborhood relocation among youth in vulnerable families: Results from a randomized trial. Archives of General Psychiatry. 2012;69(12):1284-94.

27. Vitaro F, Brendgen M, Giguere C-E, Tremblay RE. Early prevention of life-course personal and property violence: A 19-year follow-up of the Montreal Longitudinal-Experimental Study (MLES). Journal of Experimental Criminology. 2013;9(4):411-27.

28. Altman DG, Bland JM. How to obtain the confidence interval from a P value. BMJ : British Medical Journal. 2011;343.

29. Altman DG, Bland JM. How to obtain the P value from a confidence interval. BMJ. 2011;343.

30. Hackshaw A. Statistical Formulae for Calculating Some 95% Confidence Intervals.

31. StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.

32. Dunn G, Maracy M, Tomenson B. Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: The role of instrumental variable methods. Statistical Methods in Medical Research. 2005;14(4):369-95.

33. McLachlan GJ, Krishnan T. The EM Algorithm and Extensions. The EM Algorithm and Extensions: John Wiley & Sons, Inc.; 2007.

34. Gilks WR, Richardson S, Spiegelhalter D. Markov Chain Monte Carlo in Practice: CHAPMAN & HALL/CRC; 1995. 504 p.

35. Frangakis CE, Rubin DB. Principal Stratification in Causal Inference. Biometrics. 2002;58(1):21-9.

36. Rubin DB. Bayesian Inference for Causal Effects: The Role of Randomization. The Annals of Statistics. 1978;6(1):34-58.

37. Basu D. Randomization Analysis of Experimental Data: The Fisher Randomization Test. Journal of the American Statistical Association. 1980;75(371):575-82.

38. Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with noncompliance. 1997:305-27.

39. Little RJ, Yau L. Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. Psychological Methods. 1998;3(2):147-59.

40. Hirano K, Imbens GW, Rubin DB, Zhou XH. Assessing the effect of an influenza vaccine in an encouragement design. Biostatistics. 2000;1(1):69-88.

41. Jo B. Estimation of intervention effects with noncompliance: Alternative model specifications. Journal of Educational and Behavioral Statistics. 2002;27(4):385-409.

42. Frangakis CE, Rubin DB, Zhou XH. Clustered encouragement designs with individual noncompliance: bayesian inference with randomization, and application to advance directive forms. Biostatistics. 2002;3(2):147-64.

43. Jo B, Asparouhov T, Muthen BO. Intention-to-treat analysis in cluster randomized trials with noncompliance. Statistics in medicine. 2008;27(27):5565-77.

44. Jo B, Asparouhov T, Muthén BO, Ialongo NS, Brown CH. Cluster Randomized Trials With Treatment Noncompliance. Psychological methods. 2008;13(1):1-18.

45. Lin JY, Have TRT, Elliott MR. Longitudinal Nested Compliance Class Model in the Presence of Time-Varying Noncompliance. Journal of the American Statistical Association. 2008;103(482):462-73.

46. Cheng J, Small DS. Bounds on Causal Effects in Three-Arm Trials with Non-Compliance. Journal of the Royal Statistical Society Series B (Statistical Methodology). 2006;68(5):815-36.

47. Brumback BA, He Z, Prasad M, Freeman MC, Rheingans R. Using structural-nested models to estimate the effect of cluster-level adherence on individual-level outcomes with a three-armed cluster-randomized trial. Statistics in Medicine. 2014;33(9):1490-502.

48. Long Q, Little RJ, Lin X. Estimating Causal Effects in Trials Involving Multi-Treatment Arms Subject to Non-compliance: A Bayesian framework. Journal of the Royal Statistical Society Series C, Applied statistics. 2010;59(3):513-31.

49. Yau L, Little RJ. Inference for the complier-average causal effect for longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. Journal of the American Statistical Association. 2001;96(456):1232-44.

50. Huang S, Cordova D, Estrada Y, et al. An application of the Complier Average Causal Effect analysis to examine the effects of a family intervention in reducing illicit drug use among high-risk Hispanic adolescents. Family process. 2014;53(2):336-47.

51. Mealli F, Imbens GW, Ferro S, Biggeri A. Analyzing a randomized trial on breast self-examination with noncompliance and missing outcomes. Biostatistics (Oxford, England). 2004;5(2):207-22.

52. Shrier I, Steele RJ, Verhagen E, et al. Beyond intention to treat: What is the right question? Clinical Trials. 2014;11(1):28-37.

53. Jo B. Model misspecification sensitivity analysis in estimating causal effects of interventions with non-compliance. Statistics in Medicine. 2002;21(21):3161-81.

54. Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomized clinical trials. Statistics in Medicine [Internet]. 1997; 16(9):[1017-29 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcmr/articles/CMR-2411/frame.html.

55. Matsui S. Stratified analysis in randomized trials with noncompliance. Biometrics. 2005;61(3):816-23+95.

56. Little RJ, Long Q, Lin X. A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. Biometrics. 2009;65(2):640-9.

57. Nagelkerke N, Fidler V, Bernsen R, Borgdorff M. Estimating treatment effects in randomized clinical trials in the presence of non-compliance. Statistics in Medicine. 2000;19(14):1849-64.

58. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics - Theory and Methods. 1991;20(8):2609-31.

59. Korhonen PA, Laird NM, Palmgren J. Correcting for non-compliance in randomized trials: An application to the ATBC Study. Statistics in Medicine. 1999;18(21):2879-97.

60. Ma Y, Roy J, Marcus B. Causal models for randomized trials with two active treatments and continuous compliance. Stat Med. 2011;30(19):2349-62.

61. Cox DR, Oakes D. Analysis of Survival Data. First ed: Chapman & Hall/CRC; 1984.

62. Wu Y, Zhao L, Hou Y, Li K, Zhou X. Correcting for non-compliance in randomized non-inferiority trials with active and placebo control using structural models. Statistics in Medicine. 2015;34(6):950-65.

63. Ye C, Beyene J, Browne G, Thabane L. Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study. BMJ Open. 2014;4(6).

64. Odondi L, McNamee R. Performance of statistical methods for analysing survival data in the presence of non-random compliance. Statistics in Medicine. 2010;29(29):2994-3003.

65. Loeys T, Goetghebeur E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. Biometrics. 2003;59(1):100-5.

66. Adewuyi TE, MacLennan G, Cook JA. Non-compliance with randomised allocation and missing outcome data in randomised controlled trials evaluating surgical interventions: a systematic review. BMC Research Notes. 2015;8:403.

67. O'Malley AJ, Normand SLT. Likelihood methods for treatment noncompliance and subsequent nonresponse in randomized trials. Biometrics. 2005;61(2):325-34+647.

Highlights:

What is already known

- Randomised controls trials (RCTs) often suffer from non-adherence or non-compliance of trial participants to the intervention(s) protocol to which they are randomised.
- Per-protocol (PP) and as-Treated (AT) are two naïve analytical methods for handling non-adherence in RCTs, which are prone to serious selection bias and cannot claim causal treatment effect.
- Several statistical applications based on causal inference are now available to more appropriately adjust treatment effect for non-adherence in RCTs data.

What this study adds

- Our methodological review shows that, a large proportion of RCTs continue to rely on naïve PP method for handling intervention non-adherence.
- Maximum likelihood (ML) based Complier Average Causal Effect (CACE) and Instrumental Variable (IV) methods are applied in various settings to handle non-adherence in RCTs.